

СОЗДАНИЕ АНИМИРОВАННОЙ ВИЗУАЛИЗАЦИИ ПРЕДЕЛЬНЫХ РАСПРЕДЕЛЕНИЙ НЕКОТОРЫХ СТАТИСТИЧЕСКИХ КРИТЕРИЕВ НА ЯЗЫКЕ R С ПОМОЩЬЮ ПАКЕТА ANIMATION

Пасюк М.С.

Гродненский государственный медицинский университет, Беларусь

Кафедра медицинской и биологической физики

Научный руководитель – ст. преподаватель Копыцкий А.В.

Актуальность. Одним из разделов математической статистики является раздел посвящённый проверке статистических гипотез. Известно, что при проверке статистической гипотезы определяется условная вероятность p получить наблюдаемое значение статистики критерия S , попадающее в некоторый интервал значений Ψ , в предположении справедливости нулевой гипотезы. Если полученное в результате статистического эксперимента значение вероятности p окажется меньше некоторого заранее заданного порогового значения вероятности α (например, 5% или 1%), то у исследователя будут основания считать нулевую гипотезу неверной для наблюдаемых данных и её отклонить (с вероятностью ошибки p). Также известно, что для статистических критериев существуют предельные теоретические распределения их статистик, к которым они стремятся при неограниченном числе повторных измерений объектов, взятых из неизменных генеральных совокупностей. Для многих студентов, изучающих основы статистики, является трудным представить себе абстрактное стремление статистики критерия к теоретическому пределу. Таким образом, актуальным является создание наглядного способа визуализации стремления распределения статистики критерия к теоретическому распределению.

Цель работы: создание анимаций, визуализирующих то, как распределение статистики критерия при справедливости нулевой гипотезы стремится к предельному теоретическому распределению при большом количестве повторных измерений.

Для достижения цели исследования нами были решены следующие задачи:

1. Была создана генеральная совокупность из 1 000 000 чисел, распределённых по нормальному стандартному закону.
2. Для статистических критериев: Т-критерия Уэлча, F-критерия Фишера, U-критерия Манна – Уитни были проведены 150 000 повторных извлечений выборок из генеральной совокупности, и при каждом извлечении были вычислены значения статистик.
3. Для каждого массива статистик критериев были построены гистограммы их распределений. Эти гистограммы были смонтированы с графические файлы с расширением «*.gif», а также в видеофайлы с расширением «*.mp4».

Методы исследования. Для решения описанных выше задач нами был использован язык программирования «R» и интегрированная среда разработки «RStudio» [1] с подключенным пакетом «animation».

Результаты. Итак, в ходе решения задач нашего исследования были получены 6 визуализаций (по 2 для разных объёмов выборок на каждый критерий), как в графическом, так и в видео-формате. С их помощью наглядно демонстрируется, что с увеличением количества повторных измерений гистограмма распределения статистики критерия стремится по форме к соответствующему теоретическому распределению, и чем больше объёмы выборок, тем быстрее этот процесс.

Выводы. Полученные нами файлы могут быть наглядными пособиями, показывающими, что распределение статистики критерия стремится к предельному теоретическому распределению.

Литература:

1. [Электронный ресурс]. – URL: <https://www.rstudio.com/products/rstudio-desktop/> / (дата обращения 01.11.2016)